DOCUMENT RESUME

ED 235 202 TM 830 615

AUTHOR Herring, Richard D.
TITLE Finding "Problem Types" in Judgments of Problem-Similarity: Comparison of Cluster Analysis with Subject Protocols.
PUB DATE Apr 81
NOTE 13p.; Paper presented at the Annual Meeting of the American Educational Research Association 65th, Los Angeles, CA, April 13-17, 1981).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Algebra; Classification; *Cluster Analysis; Content Analysis; *Evaluative Thinking; Higher Education; Mathematics; *Memory; *Problem Solving

ABSTRACT

Literature in mathematic problem-solving suggests that learners store information in memory which helps them solve stereotyped algebra word problems. Cluster analysis has been used as an exploratory tool to infer the types of problems which have common representations in memory. This study compares the results of a hierarchical cluster analysis of judgments of problem similarity, with the open-ended rationales that subjects give for their similarity judgments. Results indicate that interpretation of a hierarchical cluster analysis is highly dependent upon the analyst's criterion of similarity. Evidence in support of two conflicting hypotheses can be taken from the same output. A critical assumption of the Johnson method of hierarchical clustering is that the data satisfy the "ultrametric inequality," and it is found that this assumption is more stringent than is commonly realized. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED235202

Finding "Problem Types" in Judgments of Problem - Similarity: Comparison of Cluster Analysis with Subject Protocols

Richard D. Herring

Tufts University

Medford, MA 02155

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

R.D. Herring

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ABSTRACT

Literature in mathematic problem-solving suggests that learners store information in memory which helps them solve stereotyped algebra word problems. Cluster analysis has been used as an exploratory tool to infer the types of problems which have common representations in memory. This study compares the results of a hierarchical cluster analysis of judgments of problem similarity, with the open-ended rationales that subjects give for their similarity judgments. Results indicate that interpretation of a hierarchical cluster analysis is highly dependent upon the analyst's criterion of similarity. Evidence in support of two conflicting hypotheses can be taken from the same output. A critical assumption of the Johnson method of hierarchical clustering is that the data satisfy the "ultrametric inequality," and it is found that this assumption is more stringent than is commonly realized.

TM 830 615

Finding "Problem Types" in Judgments of Problem - Similarity: Comparison of Cluster Analysis with Subject Protocols

Richard D. Herring

Tufts University

Medford, MA 02155

Abstract

Literature in mathematic problem-solving suggests that learners store information in memory which helps them solve stereotyped algebra word problems. Cluster analysis has been used as an exploratory tool to infer the types of problems which have common representations in memory. This study compares the results of a hierarchical cluster analysis of judgments of problem similarity with the open-ended rationales that subjects give for their similarity judgments. Results indicate that interpretation of a hierarchical cluster analysis is highly dependent upon the analyst's criterion of similarity. Evidence in support of two conflicting hypotheses can be taken from the same output. A critical assumption of the Johnson method of hierarchical clustering is that the data satisfy the "ultrametric inequality," and it is found that this assumption is more stringent than is commonly realized.

Recent investigations in mathematical and technical problem-solving have proposed the possibility that when individuals classify problems into categories, they classify these problems upon the basis of the solution methods needed to solve them, (Herrmann & Schoenfeld, Note 1; Hinsley, Hayes, & Simon, 1977; Kulm & DeGuire, Note 2; Silver, 1979, Note 3). Additional studies suggest that experts and novices differ with respect to rationales offered in support of classification schemes for solution of technical problems. Experts classify problems upon the basis of solution methods needed to solve them, novices categorize the problems upon the basis of the

descriptions of the problems, or the similarity of items described in the problems (Chi, Feltovich, & Glaser, Note 4; Larkin, McDermott, Simon, & Simon, 1980).

The experts could be characterized as classifying problems upon the basis of "deep structure" cues which are implicit in the problem statements, and the novices could be characterized as classifying them upon the basis of "surface structure" cues which are explicit in the problem statements. Cluster analysis is a tool which has been used to analyze such classification data to identify the bases for subjects' categorizations. Its use is based on the assumption that distinct categories of problems exist in subjects' memories, and that subjects will sort problems which are members of these categories into separate groups.

The present experiment is in part an attempt to replicate the results of Hinsley et al. (1977) which suggest that subjects classify algebra word problems upon the basis of deep structure cues by grouping together problems which share similar solution methods. Individuals were asked to classify problems in a card-sorting task, and to later provide rationales for their classifications. A cluster analysis of the sorting data, and interpretation of the rationales were used to assess whether individuals based their classifications on surface or deep structure cues in the problems. Limitations of the method of hierarchical clustering are discussed.

Method

Problems. Seventy-four algebra word problems were selected from a standard text on college algebra (Dolciani, Berman, & Wooton, 1970). Three problems were selected to represent each of the eighteen categories of problems identified by Hinsley et al. (1977) in their cluster analysis. The remaining 20 miscellaneous problems did not appear to fit any of these categories.

Subjects. Subjects were 19 volunteers from the Department of Psychology Research Participant Pool. They were tested in groups of four or five students at one time.

Procedure. The subjects were asked to sort the 74 algebra word problems into piles of problems which they "would solve in the same way." The problems were typed one to a page, and shuffled to provide a random order for each subject. Subjects were allowed to use a miscellaneous category, and to re-sort problems if they wished. After completing the task, the subjects listed the problems which they placed in each pile, and described how they judged the problems within each pile to be similar.

Data Analysis. The judgment data for all subjects were combined in a 74 x 74 item similarity matrix. Each entry in this matrix, A_{ij}, equaled the number of individuals who had paired problems i and j into the same category in their sort. The matrix was subjected to a hierarchical clustering technique (Johnson, 1967; diameter method).

The experimenter read the rationales the subjects gave for forming their categories, and provided a one- or two-word descriptor summarizing each. Approximately 50 different rationales were suggested in the subjects' protocols, and they are listed in Table 1.

Results

The mean number of categories formed by each subject was 10, and the standard deviation of this mean was 4.5.

The results from the cluster analysis were interpretable as problem categories, but these clusters did not conform to those reported by Hinsley et al. (1977). Both Hinsley et al. (1977) and the present analysis report the clusters formed when at least 50% of the subjects judge all problems within a set as similar. The clusters identified by Hinsley et al. and the

current study are listed in Table 2 and Table 3 respectively.

Our results indicate that interpretation of a cluster sort is highly dependent upon the criterion of similarity which the data analyst adopts. In our cluster solution, if a high criterion of similarity is adopted, few of the algebra problems are contained in the solution, but the small clusters which result are based upon method of solution which is a deep structure cue. For example, if adopting a criterion of 16 of 19 (85%) subjects judging problems as similar, seven clusters are formed which contain only 16 of the 74 problems, and the clusters appear to be based upon similarity of solution method, (river steamer, probability, interest-1, interest-2, interest-work, geometrical progressions, and area problems). If a low criterion of similarity is adopted, most problems are contained in the solution, and the large clusters which result appear based upon the contextual features which are part of the problem statements which are surface cues. For example, adopting a criterion of 2 of 19 subjects (11%) judging problems as similar, the eight clusters formed contain all 74 problems, and the clusters are based upon the quantities that are manipulated as part of the problem statements, (fluid, moving objects, physics, probability and statistics, work, numbers, finance, and geometry problems).

The category descriptions written by the subjects suggest that subjects used a large number of criteria in judging problems as similar.

Discussion

An analysis of the cluster sort reveals that reporting clusters obtained with only a single criterion of similarity may be misleading because clusters formed at separate criteria of similarity can yield conflicting interpretations. In this study, a high criterion of similarity represents a consensus among subjects, and an interpretation of these clusters suggests that subjects

judge problems as similar upon the basis of common solution methods, or deep structure cues. A low criterion of similarity yields clusters of problems related by common surface cues; however, a low criterion of similarity represents little consensus, and seems less appropriate than a high criterion of similarity. Because of alternative interpretations of hierarchical clusterings, it is advisable to report the full cluster tree or "dendogram" to readers.

A cluster analysis of group data may obscure differences in the way individuals classify objects. If the rationale behind subjects' categories is not obvious from their sorts, an additional method of collecting data may be useful in determining subjects rationales.

A difficulty with this classification task is that the solution methods of the problems may be correlated with other identifying aspects of the problems. Previous studies have identified four or more dimensions upon which subjects classify such problems such as (a) question length, (b) what the question "ask for," (c) the use of particular quantities or units in the problem, and (d) how the questions are phrased (Kulm & DeGuire, Note 2; Silver, 1979). Because these aspects of problems can be confounded, problem statements which orthogonally vary surface and deep structure cues are useful in determining which cues subjects attend to (Herring, Note 5; Hinsley et al., 1977).

Technical Notes

Hierarchical clustering methods, such as that of Johnson (1967), are very rapid and efficient ways to identify clusters from judgment data, but they have several disadvantages. One of these disadvantages is that the

items which are clustered are members of subsets which are nested in a hierarchical family, and no item can be in more than one subset, (overlapping subsets are not permitted). It is likely that items placed in separate subsets by the clustering scheme may be similar to each other because they share common features, but this similarity can not be represented because the subsets are mutually exclusive. This limitation does not exist in nonhierarchical clustering schemes (Shepard & Arabie, 1979), but operation of these schemes is often time-consuming and costly.

Hierarchical cluster techniques often presume that the data satisfy a metric axiom called the "ultrametric inequality," and this is a very restrictive assumption. In formal terms, the similarity measures between any two items can be represented as a distance function, $d(x,y)$, and any three distances must satisfy the following inequality:

$$d(x,z) \leq \max (d(x,y), d(y,z))$$

This inequality implies the following about the interpoint distances between the points A, B, C.

$$d(A,B) \leq d(A,C) = d(B,C).$$

If the distance relations between any three points are interpreted geometrically, the interpoint distances of A, B, C can only be represented as an equilateral triangle, or an isosceles triangle in which the interpoint distance between A and B, $d(A,B)$, is represented as the base, and this distance is less than the distances represented by the remaining legs of the triangle (Shepard, 1980).

The ultrametric inequality is violated under a number of conditions, including the case in which each of the three distances between three points is unique. If these interpoint distances are presumed to be related to psychological similarities, then the ultrametric inequality provides an

unreasonable assumption for most sets of psychological similarities, and for most sets of stimuli which produce these similarities.

Reference Notes

1. Herrmann, D., & Schoenfeld, A. Changes in perceptions of relatedness among mathematics problems with learning. Paper presented at the meeting of the Psychonomics Society, St. Louis, MO., November, 1980.
2. Kulm, G., & DeGuire, L. Memory structures of mathematical problems. Paper presented at the meeting of the American Educational Research Association, Boston, MA., April, 1980.
3. Silver, E. A. Problem-solving performance and perceptions of problem similarity: Retrospect and prospect. Paper presented at the meeting of the American Educational Research Association, Boston, MA., April, 1980.
4. Chi, M. T. H., Feltovich, P. J., & Glaser, R. Representation of physics knowledge by experts and novices (Technical Report No. 2). Pittsburgh, PA.: Research and Development Center, University of Pittsburgh, March, 1980.
5. Herring, R. D. Solving algebra word problems: Comparing two instructional strategies. Unpublished manuscript, University of Iowa, 1980.

References

Dolciani, M. P., Berman, S., & Wooton, W. Book two - Modern algebra and trigonometry: Structure and method. Boston, MA: Houghton Mifflin, Co., 1970.

Hinsley, D.A., Hayes, J. R., & Simon, H. A. From words to equations: Meaning and representation in algebra word problems. In P. A. Carpenter & M. A. Just (Eds.), Cognitive processes in comprehension. Hillsdale, NJ: Lawrence Erlbaum Associates, 1977.

Johnson, S. C. Hierarchical clustering schemes. Psychometrika, 1967, 32, 241-254.

Larkin, J., McDermott, H. J., Simon, D. P., & Simon, H. A. Expert and novice performance in solving physics problems. Science, 1980, 208, 1335-1342.

Shepard, R. N. Multidimensional scaling, tree-fitting, and clustering. Science, 1980, 210, 390-398.

Shepard, R. N., & Arabie, P. Additive clustering: Representation of similarities as combinations of discrete overlapping properties. Psychological Review, 1979, 86, 87-123.

Silver, E. A. Student perceptions of relatedness among mathematical verbal probelms. Journal for Research in Mathematics Education, 1979, 10, 195-210.

Table 1 - Pilot Experiment 1

Rationales which subjects listed for performing their problem classification

1. Scales
2. Physics or Chemistry
3. Geometry
4. D=RT
5. Perimeter
6. Progressions
7. "Overtake"
8. Probability
9. Areas, capacity
10. Cost & Profit
11. Velocity
12. Equations using exponents
13. Angles
14. Bearing
15. Rates
16. Ratio
17. Permutation, combination
18. Interest
19. Percentages
20. Maximum-minimum
21. When two people moving at different rates meet
22. Time-Distance
23. Accounting
24. Ratios and volumes
25. Growth and increase overtime
26. Geometric and arithmetic sequence
27. Bearing and current-type
28. Algebra and geometry
29. Calculus
30. How far, how long, what is, what were (13 types of interrogative questions)
31. Arithmetic plug-ins
32. Time, rate, speed, distance, height
33. Mixture
34. Electrical
35. Algebra number
36. Profit
37. Travel
38. "Real life" problems
39. "Worthless" problems
40. High-school type problems
41. Business
42. Common statistics
43. Quantitative or math skill
44. Work
45. Simultaneous equations
46. Applied and everyday problems
47. More difficult than applied problems
48. "Questions I read and re-read"
49. "Questions I read 1 or 2 times"

Table 1- Hinsley, Hayes, and Simon (1977)

Representative Problem from each of 18 clusters

N = 14

Similarity criterion = 7 out of 14

1. Triangle
2. DRT
3. Averages
4. Scale Conversion
5. Ratio
6. Interest
7. Area
8. Max-min
9. Mixture
10. River current
11. Probability
12. Number
13. Work
14. Nagivation
15. Progressions
16. Progressions - 2
17. Physics
18. Exponentials

Total # of problems in clusters = 64 out of 76

Table 3 - Pilot Experiment 1

Representative label for each of 17 clusters

HICLUS solution - diameter method

N = 19

Similarity criterion = 10 out of 19

1. Liquid - Mixture and exponential
2. Liquid ratio
3. Bearing & speed
4. Bearing & angle
5. Falling body
6. Variants of D=RT
7. Miscellaneous - 1
8. River, stream, & complex D=RT
9. Probability
10. Permutation
11. Combines work & work rate
12. % profit
13. Interest & compound interest
14. Interest variants - work-type
15. Arithmetic & geometric progressions
16. Miscellaneous - 2
17. Area & geometry

Total # of problems in clusters = 54 out of 74